Sentiment analysis of depression related discussions in the Suomi24 discussion forum

Jonas Tana
Arcada University of Applied Sciences
jonas.tana@arcada.fi
https://orcid.org/0000-0002-6821-1571

Andrey Shcherbakov
Arcada University of Applied Sciences
andrey.shcherbakov@arcada.fi

Leonardo Espinosa-Leal
Arcada University of Applied Sciences
leonardo.espinosaleal@arcada.fi
https://orcid.org/0000-0001-6861-8024

Keywords: depression, information behavior, sentiment analysis, social media, temporality

How to cite: Tana, J., Shcherbakov, A., & Espinosa-Leal, L. (2022). Sentiment analysis of depression related discussions in the Suomi24 discussion forum. *Informaatiotutkimus, 41*(2–3), 157–162. https://doi.org/10.23978/inf.122663

The increase in the amount of online health information, as well as the rise search engines and social media has changed the way people access and engage with health information (Tana, 2019). The internet has become a leading platform where individuals express their concerns, thoughts, mood and feelings. These billions of digital footprints from nearly all parts of the world provide a powerful opportunity to expand the evidence base across online health information behaviour. This is particularly useful when studying health topics where traditional data is limited, or when studying stigmatizing and sensitive topics where those suffering from it are disinclined to seek professional help (Tana, 2019; Wolohan et al., 2018). One such health topic is depression, which is one of the leading global disease burdens and causes disability that affects 4.4% of the global population (WHO, 2017). Finland has the highest estimated incidence rate of mental health problems in the European Union as it has been reported that 20% of the population are affected (OECD/EU, 2018), and that 4%–9% of the population suffer from a major depression (Gonzales-Inca et al., 2022). Therefore, there is a clear need to utilize online data in combination with novel methods to broaden our knowledge about depression.

Natural Language Processing (NLP) is a subfield of artificial intelligence, that enables machines to comprehend the meaning of textual input. Sentiment analysis again, is a method based on NLP to systematically identify, extract, quantify, and study affective states of large sets of textual data, to determine and classify whether a block of text is positive, negative, or neutral. The use of language in the context of mental health is nothing new. As language often reflects how people think, professionals, especially psychiatrists, have used language as a tool to assess individuals mental health conditions (Wolohan, 2018). The use of NLP however, provides the opportunity to analyse large amounts of textual data with the help of machine learning algorithms. A majority of the research conducted within this field has utilized different social media, including discussion forums, as a data source. However, the use of non-English language datasets is still scarce, thus providing us with limited knowledge about local and regional depression related online behaviours and their contents (Ibrahim et al., 2019; Zhang et al., 2022). We have set out to employ NLP based sentiment analysis on the largest Finnish anonymous discussion forum, Suomi24, to gain deeper knowledge on the emotional contents of depression related online discussions in Finland. The aim of our preliminary research is to identify temporal variations and changes in sentiment in depression related discussions.

Data used in this preliminary study was obtained from the Suomi24 discussion forum, for the whole year 2019. We filtered the data to include only the posts from the depression subcategory within the Health category (fi. Terveys > Henkinen hyvinvointi ja mielenterveys > Masennus). The total number of posts in 2019 was 3968. Sentiment analysis of the posts was performed using the Hugging Face python NLP library, including an available pretrained model for sentiment analysis of Finnish language text: finbert-finnsentiment (`https://huggingface.co/fergusq/finbert-finnsentiment`). This model is a FinBERT model fine-tuned with the FinnSentiment dataset (Lindén et al., 2020). For this preliminary study, we performed a classification into three polarities, positive, negative, and neutral, for the total number of available forum posts, separated by day and night. We use the standard separation from 06:00 to 18:00 for daytime and 18:00 to 06:00 for night-time.

The analysis shows that during daytime, posts in the depression subcategory are mostly neutral. As can be seen in Figure 1, the number of negative posts is higher than positive ones and spikes in negative sentiment are visible in February and August.

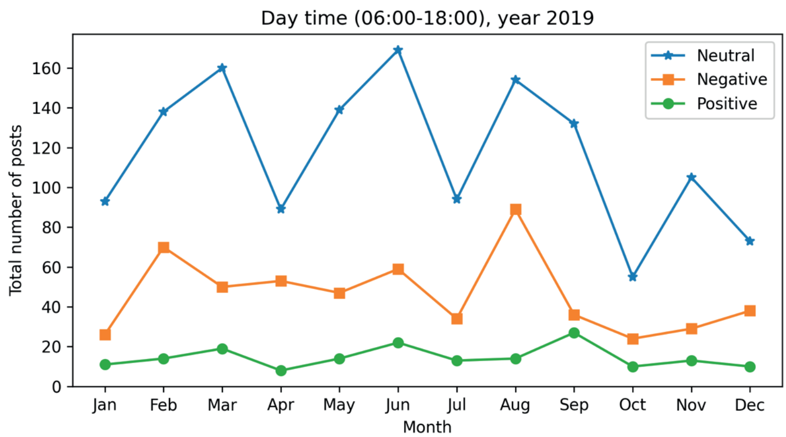

Figure 1. The sentiment categorization of positive, neutral and negative posts within the depression subcategory during daytime (06:00 – 18:00) during 2019.

During night-time, neutral sentiment posts are also dominant, with an exception in February, where the number of negative sentiment posts is slightly higher. The February and August spikes are also more distinct compared to daytime sentiment categorization.

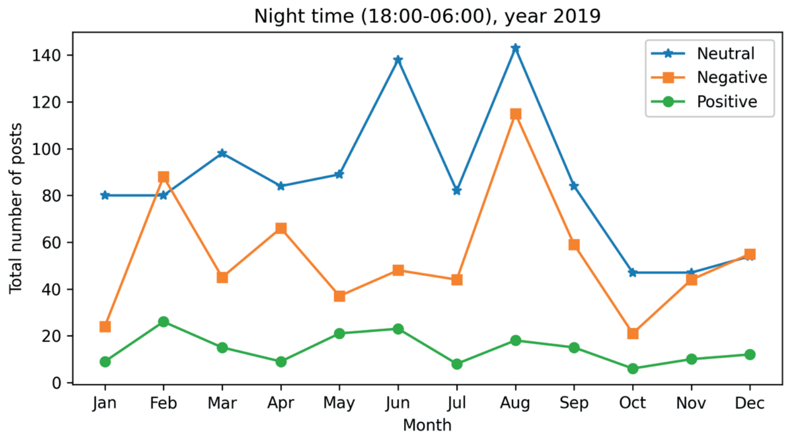

Figure 2. The sentiment categorization of positive, neutral and negative posts within the depression subcategory during night time (18:00 – 06:00) during 2019.

As can be seen in Figure 3, negative sentiment in forum posts for both day and night follow a similar pattern, with peaks coinciding. However, the relative number of negative comments is higher during night time.

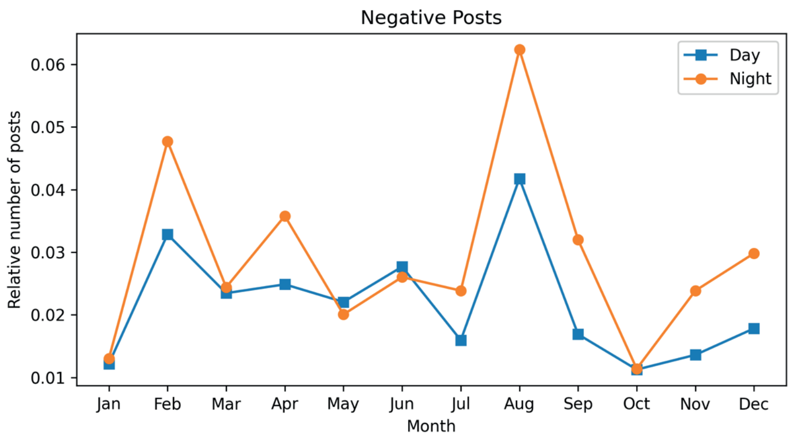

Figure 3. The normalized (over the total number of posts) monthly variation of negative posts within the depression subcategory of the Suomi24 discussion forum during the year 2019.

What our preliminary analysis shows is that spikes in posts with negative sentiment follows a bimodal seasonal pattern, with peaks in February and August. Negative sentiment is also more distinct during night time, which in previous research has been shown to be the time of day when depression related discussions in Suomi24 is more active (Tana et al., 2020). The February and August peaks in negative sentiment again are interesting, as this is a time when discussion activity in the Suomi24 depression subcategory has been shown to be lower (Tana et al., 2020). This preliminary analysis is limited to one year, and only studies the seasonal variations of sentiment. For future research, we will broaden the scope, and include discussion forum data from 2001 to 2021, as well as different timescales, to provide a more thorough analysis.

In general, the amount of behavioural data and online engagement in relation to depression will only continue to accumulate, which results in more and more user generated online health information data. With recent advances in methods of artificial intelligence, such as NLP and sentiment analysis, possibilities to utilize these large datasets more efficiently have emerged. The exploration of medical and health-related text with NLP, and in particular mental illness detection, such as depression, has recently shown an upward trend, suggesting that there is great research value and benefit in pursuing these kinds of actions (Zhang et al., 2022). As our preliminary results show, sentiment analysis can be used to identify changes in sentiment of information behaviour over time. Not only can this serve to observe the emotional landscape of depression related online discussions, but also help develop and design interventions, support as well as public health campaigns in temporally relevant fashion.

References

Gonzales-Inca, C., Pentti, J., Stenholm, S., Suominen, S., Vahtera, J., & Käyhkö, N. (2022). Residential greenness and risks of depression: Longitudinal associations with different greenness indicators and spatial scales in a Finnish population cohort. *Health & Place, 74*, 102760.

Ibrahim, M., Eteläperä, M., Turkmen, S., Maged, M., Oussalah, M., & Miettunen, J. (2019, December). Mining Health Discussions on Suomi24. In *2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking* (pp. 1580–1585). https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00232

Lindén, K., Jauhiainen, T., & Hardwick, S. (2020). FinnSentiment--A Finnish Social Media Corpus for Sentiment Polarity Annotation. *arXiv preprint*, arXiv:2012.02613.

Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of google trends in health care research: a systematic review. *PloS one, 9*(10), e109583. https://doi.org/10.1371/journal.pone.0109583

OECD/European Union (2018). *Health at a Glance: Europe 2018: State of Health in the EU Cycle*. OECD Publishing. https://doi.org/10.1787/health_glance_eur-2018-en

Tana, J. (2019). *Infodemiology: Studying rhythmicity in online health information behaviour* [Doctoral dissertation, Åbo Akademi University].

Tana, J., Eirola, E., & Eriksson-Backa, K. (2020). Exploring temporal variations of depression-related health information behaviour in a discussion forum: The case of Suomi24. *Information Research, 25*(2), paper 854. http://informationr.net/ir/25-2/paper854.html

WHO (2017). Depression and Other Common Mental Disorders: Global Health Estimates, Geneva. Available at: https://www.who.int/publications/i/item/depression-global-health-estimates

Wolohan, J. T., Hiraga, M., Mukherjee, A., Sayyed, Z. A., & Millard, M. (2018). Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with NLP. In *Proceedings of the first international workshop on language cognition and computational models* (pp. 11–21). Association for Computational Linguistics.

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. *NPJ digital medicine, 5*(1), 1–13. https://doi.org/10.1038/s41746-022-00589-7